5,837,250

GCACGTCGCATGGAGACCACCGTGAACGCCCACCAAATAT

**** ****** ***************** *

TGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCTCAGC

<u>HNF4</u>

****

<u>AATGTCAACGA</u>CCGACCTTGAGGCATACTTCAAA<u>GACTGT</u>

<u>HNF3-1</u>   HNF

* ***** ** ********** **************

<u>TTGTTTAAAGACTGGGGAGGAGTTGGGGGAGGAGATTAGGT</u>

3-2

*** * * *******

TAAAGGTCTTTGTACTAGGAGGCTGTGTAGGCATAAATTGGT

***

CTGGCGCACCAGCACCATGCAACTTTTTCACCTCTGCCTAA

Pre-genomic

******* ***

TCATCTCTTG

* nucleotide conserved at >95% among 75 HBV strains

Fig. 1A

2701 TTATTATCCAGAACATCTA<u>GTTAATCATTACT</u>TCCAAACTAGA<u>CACTATTTACAC</u>ACTCT
                        HNF1   HNF3

2761 ATGG<u>AAGGCGGGTA</u>TATT<u>TATATAA</u>GAGAGAAACAACACATAGCGCCTCATTTTGTGGGTC
      Sp1   TBP   RNA Start

2821 ACCATATTCTTGGGAACAAGATCTACAGC<u>ATGGGGC</u>
                                  PreS1 protein start
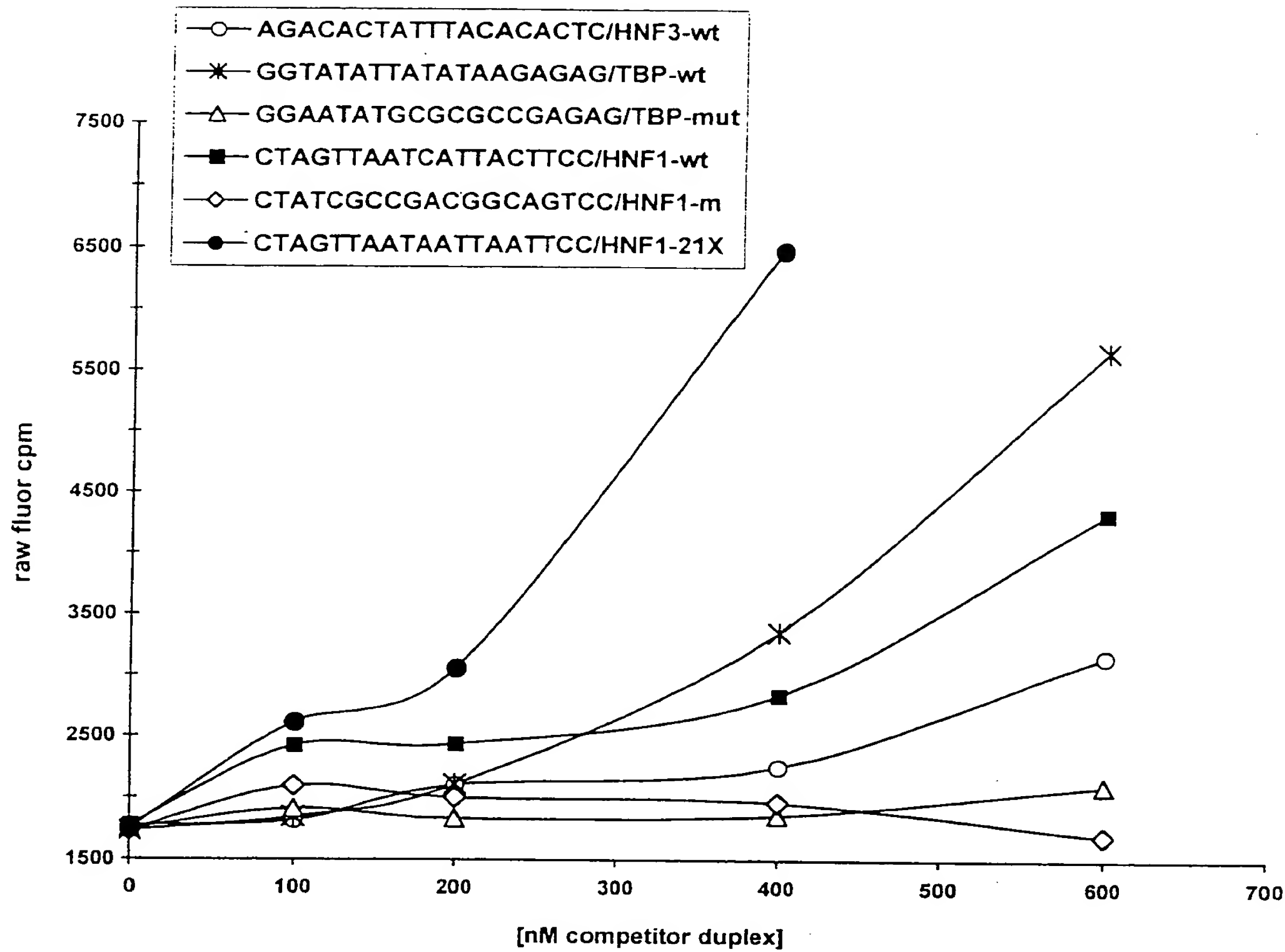
Fig. 1B

Fig. 2

1081 CTA AGC AGG CTT TCA CTT T<u>CT CGC CAA CTT ACA AGG CCT</u> TTC TGT GTA AAC AAT
      NF1(1100-1119)
                                                 2c (1119-1134)

1135 ACC TGA ACC TTT A<u>CC CCG TTG CCC GGC AAC GGC</u> CAG GTC TGT GCC <u>AAG TGT TTG</u>
      EF-C(1148-1168)

1189 <u>CTG ACG CAA CCC CCA</u> CTG <u>GCT GGG GCT TGG TCA TGG GCC ATC AGC GCA TGC GTG</u>
      E(1180-1202)   NF1(1209-1236)   X-PBP(1229-1245)

1243 <u>GAA</u> CCT TTT CGG CTC CTC TGC CGA TCC ATA CTG CGG AAC TCC TAG CCG CTT GTT

1297 TTG CTC GCA GCA GGT CTG GAG CAA ACA TTA TCG GGA CTG ATA ACT CTG TTG TCC

1351 TAT CCC GCA AAT ATA CAT CGT TTC C<u>AT GGC</u> TGC TAG   1386

Fig. 3

```
                     CAGCTGGG  CCGCCCTTGT  GCGCGGGCTG  ATGCTCTGAG  GCTTGGCTAT
          GCGGGGGCCA  ACGCGATTGT  GGGTGCTCGG  GGAGTGGGGG  GGGGCACGAC  CGTAGGTGCT
          CCCTGCTGGG  GCAACCCATC  GCTCCCCATG  CGGAATCCGG  GGGTAATTAC  CCCCCCAGGA
          CCCGGAATAT  TAGTAATCCT  AATTCCCGGC  GGGGGAGGGG  GCGCGGGAGG  AATTCACCCT
          GAAAGGTGGG  GGTGGGGGGG  GTCGCATCTT  GCTGTGAGCA  CCCTGGCGAA  GGGGAGAGGG
          CTTTTTCTAT  CAGTTTTCTT  TGAGCTTTTA  CTGTTAAGAG  GGTACGGTGG  TTTGATGACA
          CTGAACTATA  TTCAAAAGGA  AGTAAATGAA  CAGTTTTCTT  AATTTGGGGC  AGGTACTGTA
          AAAATAAAAA  CAAAAGTTAA  GACAGTAAAA  TGTCCTTTTA  TTTTTTAATG  CACCAAAGAG
          ACAGAACCTG  TAATTTTAAA  AACTGTGTAT  TTTAATTTAC  ATCTGCTTAA  GTTTGCGATA
          ATATTGGGGA  CCCTCTCATG  TAACCACGAA  CACCTATCGA  TTTTGCTAAA  AATCAGATCA
          GTACACTCGT  TTGTTTAATT  GATAATTGTT  CTGAATTATG  CCGGCTCCTG  CCAGCCCCCT
          CACGCTCACG  AATTCAGTCC  CAGGGCAAAT  TCTAAAGGTG  AAGGGACGTC  TACACCCCCA
          ACAAAACCAA  TTAGGAACTT  CGGTGGTCTT  GTCCCAGGCA  GAGGGGACTA  ATATTTCCAG
          CAATTTAATT  TCTTTTTTAA  TTAAAAAAAA  TGAGTCAGAA  TGGAGATCAC  TGTTTCTCAG
          CTTTCCATTC  AGAGGTGTGT  TTCTCCCGGT  TAAATTGCCG  GCACGGGAAG  GGAGGGGGTG
          CAGTTGGGGA  CCCCCGCAAG  GACCGACTGG  TCAAGGTAGG  AAGGCAGCCC  GAAGAGTCTC
          CAGGCTAGAA  GGACAAGATG  AAGGAAATGC  TGGCCACCAT  CTTGGGCTGC  TGCTGGAATT
          TTCGGGCATT  TATTTTATTT  TATTTTTTGA  GCGAGCGCAT  GCTAAGCTGA  AATCCCTTTA
          ACTTTTAGGG  TTACCCCCTT  GGGCATTTGC  AACGACGCCC  CTGTGCGCCG  GAATGAAACT
          TGCACAGGGG  TTGTGTGCCC  GGTCCTCCCC  GTCCTTGCAT  GCTAAATTAG  TTCTTGCAAT
          TTACACGTGT  TAATGAAAAT  GAAAGAAGAT  GCAGTCGCTG  AGATTCTTTG  GCCGTCTGTC
          CGCCCGTGGG  TGCCCTCGTG  GCGTTCTTGG  AAATGCGCCC  ATTCTGCCGG  CTTGGATATG
          GGGTGTCGCC  GCGCCCCAGT  CACCCCTTCT  CGTGGTCTCC  CCAGGCTGCG  TGCTGTGCCG
          GCCTTCCTAG  TTGTCCCCTA  CTGCAGAGCC  ACCTCCACCT  CACCCCCTAA  ATCCCGGGGG
          ACCCACTCGA  GGCGGACGGG  GCCCCCTGCA  CCCCTCTTCC  CTGGCGGGGA  GAAAGGCTGC
          AGCGGGGCGA  TTTGCATTTC  TATGAAAACC  GGACTACAGG  GGCAACTCCG  CCGCAGGGCA
          GGCGCGGCGC  CTCAGGGATG  GCTTTTGGGC  TCTGCCCCTC  GCTGCTCCCG  GCGTTTGGCG
          CCCGCGCCCC  CTCCCCCTGC  GCCCGCCCCC  GCCCCCTCC  CGCTCCCATT  CTCTGCCGGG
          CTTTGATCTT  TGCTTAACAA  CAGTAACGTC  ACACGGACTA  CAGGGGAGTT  TTGTTGAAGT
          TGCAAAGTCC  TGGAGCCTCC  AGAGGGCTGT  CGGCGCAGTA  GCAGCGAGCA  GCAGAGTCCG
          CACGCTCCGG  CGAGGGGCAG  AAGAGCGCGA  GGGAGCGCGG  GGCAGCAGAA  GCGAGAGCCG
          AGCGCGGACC  CAGCCAGGAC  CCACAGCCCT  CCCCAGCTGC  CCAGGAAGAG  CCCCA
```

Fig. 4

```
           10         20         30         40         50         60         70
GAATTCACTG GGGAGAGCAT TCAGGAAGAT GACAACAGGA TAATAGGTCA ACAGAGTAAT AGAGAGGTCG
CTTAAGTGAC CCCTCTCGTA AGTCCTTCTA CTGTTGTCCT ATTATCCAGT TGTCTCATTA TCTCTCCAGC

           80         90        100        110        120        130        140
CTAAAAATAA ACTCTAAGAA GTATTCAGCC AAAACTATTA TTGAGCTAAT AATGGTGGGA TCAATTTCAG
GATTTTTATT TGAGATTCTT CATAAGTCGG TTTTGATAAT AACTCGATTA TTACCACCCT AGTTAAAGTC

          150        160        170        180        190        200        210
GGGAATATTG TGGGCAGAAG TCAGACTGTA GGAGGCTGGG GATCAAGAAG TTGAGGCAAG GAGGTTGGAC
CCCTTATAAC ACCCGTCTTC AGTCTGACAT CCTCCGACCC CTAGTTCTTC AACTCCGTTC CTCCAACCTG

          220        230        240        250        260        270        280
AACAACTGTT TTTTCAAGTT GGTCACGTGA ACAAATCTGT GACCTTCAGC CTCCCCTCCC TCGGGTCTTG
TTGTTGACAA AAAAGTTCAA CCAGTGCACT TGTTTAGACA CTGGAAGTCG GAGGGGAGGG AGCCCAGAAC

          290        300        310        320        330        340        350
GCTGAGCTGA TTGCAGGGCC CCTGCAGCTC TGGCACTCTC AAGTTGTATA AAACTGACAG TGCAGAAGTC
CGACTCGACT AACGTCCCGG GGACGTCGAG ACCGTGAGAG TTCAACATAT TTTGACTGTC ACGTCTTCAG

          360        370        380        390        400        410        420
CTTGAGCCCA TTTTGGCTCT CATGATAATT TTCCTTCAGT GGAACTAAGG TTACTTGTCT AAGAACCAAA
GAACTCGGGT AAAACCGAGA GTACTATTAA AAGGAAGTCA CCTTGATTCC AATGAACAGA TTCTTGGTTT

          430        440        450        460        470        480        490
GCCTCTGACT TGACTGATCA AAGTTCATCA CGTGCATCGA AGCCACCTAC TTGGCAGATG TAGTGAAAAG
CGGAGACTGA ACTGACTAGT TTCAAGTAGT GCACGTAGCT TCGGTGGATG AACCGTCTAC ATCACTTTTC

          500        510        520        530        540        550        560
CTACATAGAT CTGGGCCCAG GACAGGATGC TGGGGCGTGG GAGGGGAAGA AAGCAGGTGC TAACTATATA
GATGTATCTA GACCCGGGTC CTGTCCTACG ACCCCGCACC CTCCCCTTCT TTCGTCCACG ATTGATATAT

          570        580        590        600        610        620        630
GATAGCATGC CTATCAGAGC AGTTTTTACG TTTCCTATTT GTCTCTCAAA ACAATTTTAT AGGAATCATC
CTATCGTACG GATAGTCTCG TCAAAAATGC AAAGGATAAA CAGAGAGTTT TGTTAAAATA TCCTTAGTAG

          640        650        660        670        680        690        700
AAAGCAATTT TATCATGGTT TCTAGACCAG GTTTGGATGT GAGGTAGGGA TTTCCACAGC TGCTTTTAGT
TTTCGTTAAA ATAGTACCAA AGATCTGGTC CAAACCTACA CTCCATCCCT AAAGGTGTCG ACGAAAATCA

          710        720        730        740        750        760        770
TTGAAGGAAA TCTGATAAGA TGATGCAAAA GCCCTTCAGA AATGTGTAAT CCTACACACT TCAGTGATTC
AACTTCCTTT AGACTATTCT ACTACGTTTT CGGGAAGTCT TTACACATTA GGATGTGTGA AGTCACTAAG

          780        790        800        810        820        830        840
AATTCATTGT CAAAACTTAA GGTGTTTTTA ATATTGTTAT TGTTCATTTG GTTTTTACCA ACATGTAAGG
TTAAGTAACA GTTTTGAATT CCACAAAAAT TATAACAATA ACAAGTAAAC CAAAAATGGT TGTACATTCC

          850        860        870        880        890        900        910
AGTTGGCAAT TATTTGTTAA ACTCATGTCT TAGGCTAAAT AAATTCCAAA AAATTCAGGA TGAGAATTGT
TCAACCGTTA ATAAACAATT TGAGTACAGA ATCCGATTTA TTTAAGGTTT TTTAAGTCCT ACTCTTAACA
```

Fig. 5A

```
          920        930        940        950        960        970        980
    TTATTGCTTA ACGTGTGTCA AATTTCTTCC ATGCACATCT TTATTAGATC TTCACAGCAA CCTACAGGAT
    AATAACGAAT TGCACACAGT TTAAAGAAGG TACGTGTAGA AATAATCTAG AAGTGTCGTT GGATGTCCTA

          990       1000       1010       1020       1030       1040       1050
    AAGCAAGACA GGTGCAAGTG CCTCCTTTGG GTATGAGGAA ACTGAGGTCT AAAGAGATGA AGTGATTTGC
    TTCGTTCTGT CCACGTTCAC GGAGGAAACC CATACTCCTT TGACTCCAGA TTTCTCTACT TCACTAAACG

         1060       1070       1080       1090       1100       1110       1120
    CCAAGGCTCA TAGCAATTTA TTGGTAGAGC AAAGACTAGA ATTCTCTTAA CTGCAGCCTA TTTTCCCTAT
    GGTTCCGAGT ATCGTTAAAT AACCATCTCG TTTCTGATCT TAAGAGAATT GACGTCGGAT AAAAGGGATA

         1130       1140       1150       1160       1170       1180       1190
    TCTGAACTGT TACATCAGCA TCAACAATTA TCTAATGGAT TGGAACAGTG TACACAGGCA GCTTAGCTAC
    AGACTTGACA ATGTAGTCGT AGTTGTTAAT AGATTACCTA ACCTTGTCAC ATGTGTCCGT CGAATCGATG

         1200       1210       1220       1230       1240       1250       1260
    GTCAAGTCAC GATTTTTACT TTAACTTCAA TTCCAGAGTC TTGGCCTGAT TTCCCTCAAG ACCCTACTTA
    CAGTTCAGTG CTAAAAATGA AATTGAAGTT AAGGTCTCAG AACCGGACTA AAGGGAGTTC TGGGATGAAT

         1270       1280       1290       1300       1310       1320       1330
    TCTTTGGCTT TGGAAAATTT ATTTTTCTTG CATTATCTTT CCAGCTAAAT TTTATTTAAT AACCATCAGC
    AGAAACCGAA ACCTTTTAAA TAAAAGAAC GTAATAGAAA GGTCGATTTA AAATAAATTA TTGGTAGTCG

         1340       1350       1360       1370       1380       1390       1400
    ATGCTTTTTT TGCTTTATGC CATGTAGACT TGACCTGAAA ACCTGCCAGG CTTTCATTGA GTTTAGTGAT
    TACGAAAAAA ACGAAATACG GTACATCTGA ACTGGACTTT TGGACGGTCC GAAAGTAACT CAAATCACTA

         1410       1420       1430       1440       1450       1460       1470
    TAAAGAAGTA AAGTTCTGAG AAGCAATTAG TTGATGGGAC ACCAGTCATA AAATCAATCC AAACTTTTGT
    ATTTCTTCAT TTCAAGACTC TTCGTTAATC AACTACCCTG TGGTCAGTAT TTTAGTTAGG TTTGAAAACA

         1480       1490       1500       1510       1520       1530       1540
    TGACATGTGT TTCTTTCTCC ATATACCAGG TTCCCGCTTC GTATTAGTAA GATTGAAATT GAAATAAGTC
    ACTGTACACA AAGAAAGAGG TATATGGTCC AAGGGCGAAG CATAATCATT CTAACTTTAA CTTTATTCAG

         1550       1560       1570       1580       1590       1600       1610
    TATTGCTGGT GGATGAATTT GTCACTTTCC TTGAAACTGG TGAACCCAAA AAGTTAGACA GTGATAGGAA
    ATAACGACCA CCTACTTAAA CAGTGAAAGG AACTTTGACC ACTTGGGTTT TTCAATCTGT CACTATCCTT

         1620       1630       1640       1650       1660       1670       1680
    AATACTGCCA TTGTCTGTTA AGAAGTCTAT GACATTTCAA GGCAAGAATG AATATATGGA AGAAGAAACT
    TTATGACGGT AACAGACAAT TCTTCAGATA CTGTAAAGTT CCGTTCTTAC TTATATACCT TCTTCTTTGA

         1690       1700       1710       1720       1730       1740       1750
    TGTTTCTTCT TTACTTACAA AAAGGAAAGC CTGGAAGTGA ATGATATGGG TATAATTAAA AAAAAAAAAA
    ACAAAGAAGA AATGAATGTT TTTCCTTTCG GACCTTCACT TACTATACCC ATATTAATTT TTTTTTTTTT

         1760       1770       1780       1790       1800       1810       1820
    AAAACAAAAA ACCTTTACGT AACGTTTTGC TGGGAGAGAA GACTACGAAG CACATTTTCC AGGAAGTGTG
    TTTTGTTTTT TGGAAATGCA TTGCAAAACG ACCCTCTCTT CTGATGCTTC GTGTAAAAGG TCCTTCACAC
```

Fig. 5B

```
          1830       1840       1850       1860       1870       1880       1890
GGCTGCAACG ATTGTGCGCT CTTAACTAAT CCTGAGTAAG GTGGCCACTT TGACAGTCTT CTCATGCTGC
CCGACGTTGC TAACACGCGA GAATTGATTA GGACTCATTC CACCGGTGAA ACTGTCAGAA GAGTACGACG


          1900       1910       1920       1930       1940       1950       1960
CTCTGCCACC TTCTCTGCCA GAAGATACCA TTTCAACTTT AACACAGCAT GATCGAAACA TACAACCAAA
GAGACGGTGG AAGAGACGGT CTTCTATGGT AAAGTTGAAA TTGTGTCGTA CTAGCTTTGT ATGTTGGTTT


          1970       1980       1990       2000       2010       2020       2030
CTTCTCCCCG ATCTGCGGCC ACTGGACTGC CCATCAGCAT GAAAATTTTT ATGTATTTAC TTACTGTTTT
GAAGAGGGGC TAGACGCCGG TGACCTGACG GGTAGTCGTA CTTTTAAAAA TACATAAATG AATGACAAAA


          2040       2050       2060       2070       2080       2090       2100
TCTTATCACC CAGATGATTG GGTCAGCACT TTTTGCTGTG TATCTTCATA GAAGGCTGGA CAAGGTAAGA
AGAATAGTGG GTCTACTAAC CCAGTCGTGA AAAACGACAC ATAGAAGTAT CTTCCGACCT GTTCCATTCT


          2110       2120       2130       2140       2150       2160       2170
TGAACCACAA GCCTTTATTA ACTAAATTTG GGGTCCTTAC TAATTCATAG GTTGGTTCTA CCCAAATGAT
ACTTGGTGTT CGGAAATAAT TGATTTAAAC CCCAGGAATG ATTAAGTATC CAACCAAGAT GGGTTTACTA


          2180       2190       2200       2210       2220       2230       2240
GGATGATGGT AGAAACCAAA TAGAAGAATG GTCTTGTGGC ATAATGTTTG TTCCCTAGTC AATGAACTCT
CCTACTACCA TCTTTGGTTT ATCTTCTTAC CAGAACACCG TATTACAAAC AAGGGATCAG TTACTTGAGA


          2250       2260       2270       2280       2290       2300       2310
CATATTCTTG TCTCTGGTTA GGATCTTGGG ATCTGGAGTC AGACTGCCTG GGCTCAAATC TTGGCTCTGC
GTATAAGAAC AGAGACCAAT CCTAGAACCC TAGACCTCAG TCTGACGGAC CCGAGTTTAG AACCGAGACG


          2320       2330       2340       2350       2360       2370       2380
CCATACCATC TCTGTTATCC TGGGGCAAGT GCCTCAGTTT CCACATCTGA GAAATGGGGA TGGTAGTGGT
GGTATGGTAG AGACAATAGG ACCCCGTTCA CGGAGTCAAA GGTGTAGACT CTTTACCCCT ACCATCACCA


          2390
GTCCATTTCA TAGAT
CAGGTAAAGT ATCTA
```

Fig. 5C

GAGATGTATATAATTTTTTAGGAAAATCTCAAGGTTATCTTTACTTTTTCTTA
GGAAATTAACAATTTAATATTAAGAAACGGCTCGTTCTTACACGGTAGACTTA
ATACCGTAAGAACGAGCCGTTTTCGTTCTTCAGAGAAAGATTTGACAAGATTA
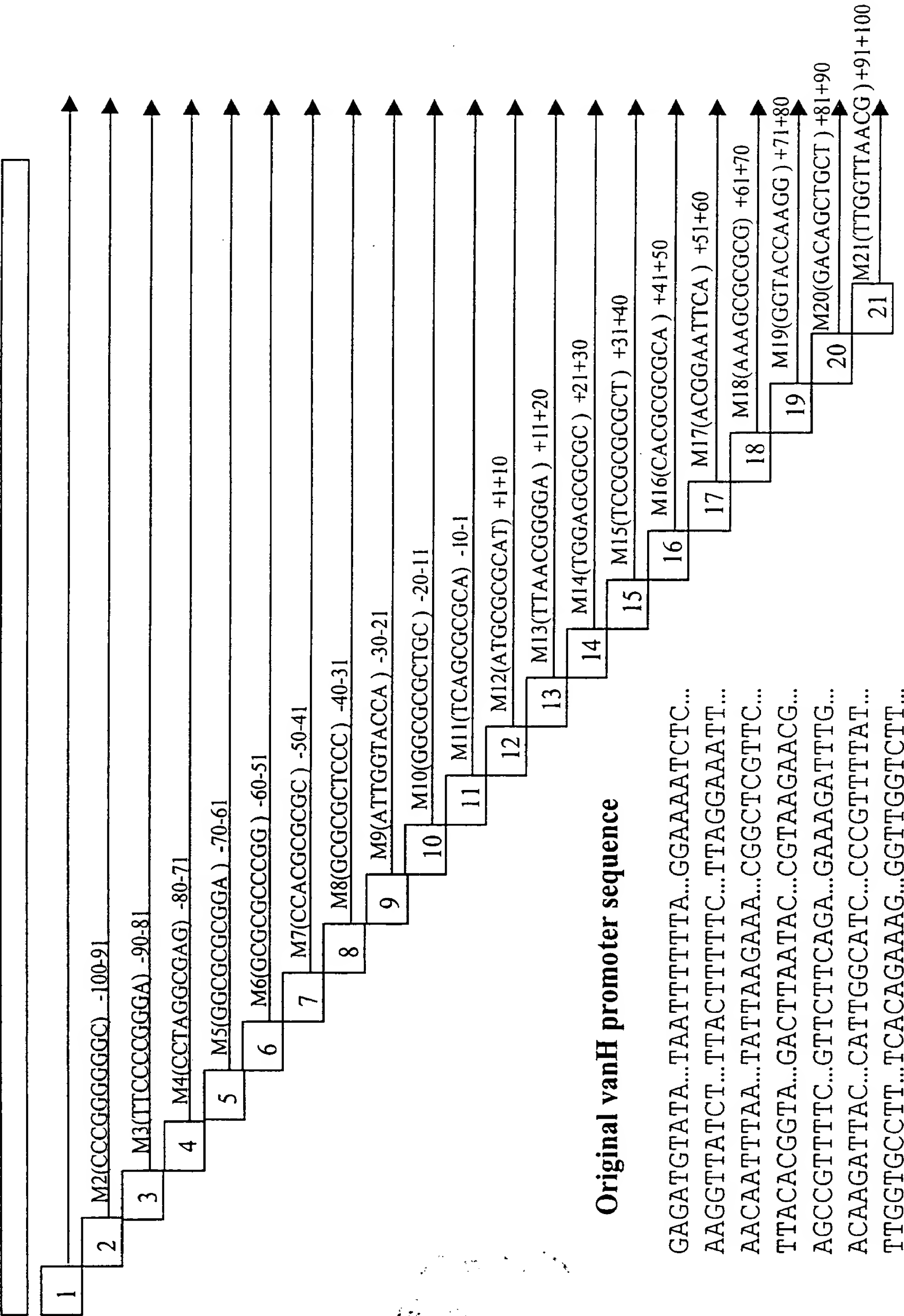CCATTGGCATCCCCGTTTTATTTGGTGCCTTTCACAGAAAGGGTTGGTCTTAA
TT

Fig. 6

Fig. 7

Original vanH promoter sequence

GAGATGTATA..TAATTTTTA..GGAAAATCTC...
AAGGTTATCT..TTACTTTTTC..TTAGGAAATT..
AACAATTAA..TATTAAGAAA..CGGCTCGTTC..
TTACACGGTA..GACTTAATAC...CGTAAGAAACG...
AGCCGTTTTC..GTTCTTCAGA...GAAAGATTTG...
ACAAGATTAC..CATTGGCATC..CCCGTTTTAT..
TTGGTGCCTT..TCACAGAAAG..GGTTGGTCTT...

M2(CCCGGGGGGGC) -100-91
M3(TTCCCCGGGA) -90-81
M4(CCTAGGCGAG) -80-71
M5(GGGCGCGCGGA) -70-61
M6(GCGCGCCCGG) -60-51
M7(CCACGCGCGC) -50-41
M8(GCGCGCTCCC) -40-31
M9(ATTGGTACCA) -30-21
M10(GGGCGCGCTGC) -20-11
M11(TTCAGCGCGCA) -10-1
M12(ATGCGCGCTGC) +1+10
M13(TTAACGGGGGA) +11+20
M14(TGGAGCGCGC) +21+30
M15(TCCGGCGCGT) +31+40
M16(CACGCGCGCA) +41+50
M17(ACGGAATTCA) +51+60
M18(AAAGGCGCGCG) +61+70
M19(GGTACCAAGG) +71+80
M20(GACAGCTGCT) +81+90
M21(TTGGTTAACG) +91+100

```
TCTAGAAAAT AATTCCCAAT ATTGAATCCC AAAGAATTCA ACATTTGGGC TGTCGTTTGA 61
AAGATAAGTT GAATTTGGTC ATGAAGGAAG AGAGGGGGGA TACAATTTCA GTAAAAGGTA 121
ACAGCAAGGT CCAAAGACAG TCAGGTCTTC AGTAGTATGG AGTATATTCA GAGGGAGCCA 181
AGATGTCTGA TGTGAACTAA AAAGATTGGT GGTTGGTAGG AGGAAGAGGT GTGAGAAGAG 241
GCTGTAAAGA AAAATTGAAA CTTGATTGTG ATGGACTTTA AAGGCTAGGC TATGGGACTT 301
GGACATGAAT CTGCAGGCCA GTGTTTGCAG ACTGGCGCCC ATAACTGTCT ATCACAGCAA 361
CACAGACATG TGTTGTTTGG CCTGCAGAGG TTTGGCCTGC ATGATGATTT TAAACCATCT 421
GAATTAGTAG CCATCATTTT CAAAAATCAA GAGATGCCAC ATTAAAATAT GGAATGCTGC 481
TGTTCTTGAA AATAATGAAA CATCTGGAAC ATTGAGGCCA CATTCCTGAC TGACAGCAAT 541
CAGTTGGAGC TGCGTAGTGA CTGCCCACTT TACATGGGGC ATCTGATCCC TAGTCGATTA 601
CAGCTGCCAC CACTTCCCTT TATCTCTCTA ATACCAAGCT CTTTTCACTC ATTTTTGTTA 661
CTTAAGAGAT ATTTGGGTTT GAAACCTCTG ATGCAGGTAA TTGAGGGTTA TAGAGCAGAG 721
GACAGATGCT ATCAGAGTTG TCTTTTAAGA AAGAACCCTC TGTTCTTCAT TTTGTTGAAG 781
ATAGCCTGGA AGAGGGCAGC CAGGGGAGAA GTTAGGGCTG GAGCTATGAG AAAGCATAAG 841
ATGAGATGAT GGCTTCAACA TTGAGGACAG AAAGAATATT GAGATGAGAA AGTAGTCCAT 901
ATAAGCATCT ATGCAAAGGA AATAGCAGAT GTCCTCAAAT CAGCAGAGGC AACAACTCTG 961
AAAGTTTATT CATAAGCCCC TCTTTTCATC TCCAATCCAG TTCAAATGTA ATTATTTAAA 1021
TTGTTCTTCA CTCTCCTTCC TGGATCATGA ATGAGCTCCT TAAATGCAGG GTCCACAGTG 1081
TCCTATTCAT CAGTGAATTC CAAGTGCCTA GCACAGAGCC TGGCAAATAG TAAATGCTTA 1141
ACAAATATTC GTTCAGTGCA TGAATTGGAG TGATTCTCTA CTTTGCCTCA TAAGTTGAAA 1201
AAAGGTTTAT TACATACCTA AATATGCTGA AATCACAGGG CATTTGGCAA CCCCCCAAAA 1261
CCAAAACTCC CAGTTTGGAA ACAGAATTTT AATTCTGTGA AAATAAAATC CATTCATTTA 1321
TTCAAAAAAT ATTTATTAAA CAATGACCAT GTCCACACCA GGCTGAGTCC TAAGGATTCA 1381
ATGATGAACA AAAACCAACA TGATTCCTGC TCTTAGGAAA CATACAGTTC AGTGAGGAAA 1441
ACAGATTGTG AGAAGTCCTC CAACAAATAC TGGGTGCTAT TAAAATATAT TAAAAGGTGA 1501
GTGGGTGAGG GACTTGAGCT AGCCTAGGTG GTTCAGGAAG TCTTCCTGGA TGTGCTGATA 1561
TGCATAGGCA TTAACTAGAT AAATAGAGAG AAGGATGAAC CAACATTGCA GGTAGAGGGA 1621
ACAGAATATG CAAAGGCAGG AAGGATTATG GAGTCGTTGG AGGACCTGAA TAAAGGCCCA 1681
GTGTAAGTGG ATCTCAGAAA ACAGGAGGAA AGGTGTATGA GATGAGATCA GAGAGGCAGA 1741
TCATGTGGGG TATGGTTAAT GTTTTGGACT TTTCTATTAA GAGCAATGGG GAGACAGTGA 1801
CAGGACTTAA ACGGGGAAAT AATATGACCA GATTAAACTT TCTAAAAAAC CCTCTATGCA 1861
AATATATATT GAGAGTTAAT TATTGACAAA GATTCAAAGG CAACAAAGTG GAGAGAGAAT 1921
AGTATTTTCA AAAAATGGTG CCAAAACAAT AGGACATCTA TATTAAAAGT TGGGTATCTG 1981
TCTACAAAAC TTAATTCAAA ATGGATCACA GACCTAAATG TAAAACTGAA AGCTATACAA 2041
CTTCTGGAAG GAAAACACAG ATGGGAATCT GTGTGATCTT GAGTTTGAAA ATGATTTATT 2101
ATATCTGACA CCATAATCCG TAAGTTAACA TAATTCATAA GTGAACAAAG TGATGAACTG 2161
GACTTCATCA GAATTTAAAA TGTTTGTGCT TCAAAGACA CTGGTATGAT AATGAAGACA 2221
AACTACAGAT AAGATATTGT TGAATCATAT TTCTGATAAA GGAATTGTGG CTCAGAATAC 2281
ATAACTCTAA ACCCCCATAA TAAATTACAA GTAGCCCAAT TAAAAAAAAA AAAAGAGAAA 2341
AAATTTACAG TCTTCATCAA AGAAAGTATC AATTGTAAAA TAAGCACATG AAAAATGCTC 2401
TGCATCTTTA TTCATGGGGG GATGAAATAA AAATTAAATG GGAAAGACAC CTCTAATTAG 2461
AATACTAAAA TTAAAAAGAC TGACCATACC AAGTATTGGT GAAGTGGAAA TGTAAAATGA 2521
TACAATCAAC TTAGGTAGAT GATTTGGAAG TTTCTTACAA AAGTAGGTGT ATACCTACCC 2581
TGTGACTCAC CCATTCCATG GCTAAGTATT TACCTGAGAG AAATGAAAGA ATACATCCAT 2641
ACAAAGATGT TTATACAAAT ATTTATAGCA GTTTTATTTG TAGTAGCCCC AAACTGAAAA 2701
GAACCCAAAT GTCCATCAAA AGTGAATGGA TAAACAAAGC GTGGTACAGC AATGCAATAG 2761
AATACTACTT AGCAATAAAG AAGAATGAGC TAGTGATATA CATAACAGCT TAAATGTACA 2821
TCAAAGGCAT TGTGCTCAGT GAAAGATGCA AGTAAAAAAA AAAAGAGTA CATGCTGTAT 2881
AGTTCCATTG ACATAAAACT CTGGAAAGTG AAAAACAGTC TATACTGACA GAAAGCAGAT 2941
CATTGGTTGC CTGAGGAGGA GGAGTATAGG AGAGGTGGAG GGAAATGTA CAAAGTGGCA 3001
CAATAAAAAC TTTTGGAATC ATAGATATAT TCACTATCTT GATTGAGTGA TGATTTCATG 3061
```

Fig. 8A

```
AGTGCACGTG CGTGTGTCAA AAATGATCAA TTTATGCAAC TTTAAATATG TGCAGTTTAT 3121
TGTATATATC AATTATACCT CAGTACGGCT ATTAAAAAGA AACCCTCTGG CTGCACAATG 3181
CAGAACTGAT TCTAGGAAAG AGTGGAGGGA GGATGACCAT TTACAGTGCT CCAGGTGGAA 3241
GAGAACGGTG CCTTCTGGAA GTGAACTAGG TTGGCAACAA CAGAGATGAA ATAAATGGGC 3301
AGATGTGTGA GATACTTAGG AAATAAAACC CGATGGTCAC CATTTTCCAA AGGTCAGCTC 3361
ATCCTGGCTT TCCAGAGCAA AGAGCTAGGG AAGACTTTAT TAATAAATCC CTCTTGAAGT 3421
TGCAGAGGAA GCTTATAGCA GAAACTTACT CTCAACCTGA CTAATCTGAG AGAACACCTC 3481
TGGTTCCATT TGATTACTAA AAAACTGCAA AGAACAGGAG GAGAAAGAAG AAGAAAGCTG 3541
GTACAAACAG TGAACTTATA TAATATTAAT CAATAATTGT CTCTTGTTCT TAAAAGCAAT 3601
GGGAAGAAAA TGAGATTTGA GCTGGAAGAT CAGAGTTCAA AATCCAAATA AAGTATATGG 3661
CCCTAATATG CTTATAGTAG TTAACCTTTC CTGATAATGA TATAATTGTT GACAGCACCA 3721
TCTTTAAAAT AAAATAACAT AGTAATCCTT CAGATTTGTA GAAGATCTTT CCTGTTTACA 3781
AGTTTGTTCT ATACACATTA TGTCTTTTAA ATGACACACT AGCCTTCTGA GGGTAACTTA 3841
TATTGGCAAC AGTTTTCAGA TGTGGAAACT GTGAAGACAA TGTTGGTGAT GTGGAAGCAA 3901
CATAAACTTT GGAGTCTTTC AGACCCAGGT TTGAATGTCA GACTGCTTTT TATTCAGAGT 3961
AACTTCAGAG CATTATTTCT CACCTTAATT TTTTTTCAGG CCTCTTTGTG TCTATGTGTC 4021
CTCTTCACTC CTGTCCATTG TTTCTTCAGT GATTTTTGCC ACCTTCCTTC ACTGTTAGTG 4081
TGTAGACACA TAGTTCTCCT GGCTCTGAGA GCCTATGTTA ATTCCATTCT ACCATCCTGC 4141
CACGGCCCAC TCAATTCCTA TTGAGCAATG CTAGTTGAAA GTTGTGGTGG GATTAAATGT 4201
TGCAATGAGT ATTCAAATGA GGTTGAAGTA TCTACGCATT CTACTTACAT ATGGTGAGGT 4261
ATATTCAAGG AAGCTGTAGC CATTAAAATC TCAGGAAATA ATTTTTCACC TCCTCAGGTG 4321
AAAGGGTCTT CAGGCCTTTG TGTTCTGGAA GGTTCATTTA TAGCCATTTC CCAAATGACA 4381
ATGCGATTGA TGAGTCTAGA GTCTAGCTCA AATAGCAATG GACTGGAAGA CTAGTTTAGG 4441
TTTTACTAAT GTGGAACATA GAACAAATTA TGTCCTTGTT TCAGCCTGTT CATCTGTGAA 4501
ATAGAGCCTA TCATATCCAG TCTTCCTTGC CTTTAGGTTT GAGTTACCTT CTTTGGTCAA 4561
GGTAAGTAAA TGCCTATGAT GTTTGGCTGT GCACAAGATA AAGCTACAAC AAAGCTACAA 4621
CCCATCTTTT CTCTGTAGAA GACTCAAAAA GCAAAAGAGA CCCAGGAAAA TCTCGGAATG 4681
ACTTTTGGAA CAGAGAGCCT CCCCAGAATC AGAAGTCAAG GAATTTAAAC ATAGGGAAGG 4741
CCCAGGTCTC TACTGACATA AAGGAAAGAT GTTTTCTTAT AGGTTTCACG TTTACATTTT 4801
CTCTCTCTTG ATCCCATTCC CACTTGCATC TGCCACCTTT ACACAGGGCT TATGGGACCT 4861
CCTCCACAAA AGAGCAGTTG CAGTAACCCA CATCATCCTC TACGCCCTGG CTGTCCATCA 4921
AGAGGCGAAA AGCAGCCCTA TATAGGTTCT ATCCTTGGAT AGTTCCAGTT GTAAAGTTTA 4981
AAATATGCGA AGGCAACTTG GAAAAGCAAG CGGCTGCATA CAAAGCAAAC GTTTACAGAG 5041
CTCTGGACAA AATTGAGCGC CTATGTGTAC ATGGCAAGTG TTTTTAGTGT TTGTGTGTTT 5101
ACCTGCTTGT CTGGGTGATT TTGCCTTTGA GAGTCTGGAG AGTAGAAGTA CTGGTTAAAG 5161
GAACTTCCAG ACAGGAAGAA GGCAGAGAAG AGGGTAGAAA TGACTCTGAT TCTTGGGGCT 5221
GAGGGTTCCT AGAGCAAATG GCACAATGCC ACGAGGCCCG ATCTATCCCT ATGACGGAAT 5281
CTAAGGTTTC AGCAAGTATC TGCTGGCTTG GTCATGGCTT GCTCCTCAGT TTGTAGGAGA 5341
CTCTCCCACT CTCCCATCTG CGCGCTCTTA TCAGTCCTGA AAAGAACCCC TGGCAGCCAG 5401
GAGCAGGTAT TCCTATCGTC CTTTTCCTCC CTCCCTCGCC CCACCCTGTT GGTTTTTTAG 5461
ATTGGGCTTT GGAACCAAAT TTCCTGAGTG CTGGCCTCCA GGAAATCTGG AGCCCTGGCG 5521
CCTAAACCTT GGTTTAGGAA ACCAGGAGCT ATTCAGGAAG CAGGGGTCCT CCAGGGCTAG 5581
AGCTAGCCTC TCCTGCCCTC GCCCACGCTG CGCCAGCACT TGTTTCTCCA AAGCCACTAG 5641
GCAGGCGTTA GCGCGCGGTG AGGGGAGGGG AGAAAAGGAA AGGGGAGGGG AGGGAAAAGG 5701
AGGTGGGAAG GCAAGGAGGC CGGCCCGGTG GGGGCGGGAC CCGACTCGCA AACTGTTGCA 5761
TTTGCTCTCC ACCTCCCAGC GCCCCTCCG AGATCCCGGG GAGCCAGCTT GCTGGGAGAG 5821
CGGGACGGTC CGGAGCAAGC CCACAGGCAG AGGAGGCGAC AGAGGGAAAA AGGGCCGAGC 5881
TAGCCGCTCC AGTGCTGTAC AGGAGCCGAA GGGACGCACC ACGCCAGCCC CAGCCCGGCT 5941
CCAGCGACAG CCAACGCCTC TTGCAGCGCG GCGGCTTCGA AGCCGCCGCC CGGAGCTGCC 6001
CTTTCCTCTT CGGTGAAGTT TTTAAAAGCT GCTAAAGACT CGGAGGAAGC AAGGAAAGTG 6061
```

Fig. 8B

```
CCTGGTAGGA CTGACGGCTG CCTTTGTCCT CCTCCTCTCC ACCCCGCCTC CCCCCACCCT 6121
GCCTTCCCCC CCTCCCCCGT CTTCTCTCCC GCAGCTGCCT CAGTCGGCTA CTCTCAGCCA 6181
ACCCCCCTCA CCACCCTTCT CCCCACCCGC CCCCCCGCCC CCGTCGCCCA GCGCTGCCAG 6241
CCCGAGTTTG CAGAGAGGTA ACTCCCTTTG GCTGCGAGCG GGCGAGCTAG CTGCACATTG 6301
CAAAGAAGGC TCTTAGGAGC CAGGCGACTG GGGAGCGGCT TCAGCACTGC AGCCACGACC 6361
CGCCTGGTTA GGCTGCACGC GGAGAGAACC CTCTGTTTTC CCCCACTCTC TCTCCACCTC 6421
CTCCTGCCTT CCCCACCCCG AGTGCGGAGC CAGAGATCAA AAGATGAAAA GGCAGTCAGG 6481
TCTTCAGTAG CCAAAAAACA AAACAAACAA AAACAAAAAA CAAGAAATAA AAGAAAAAGA 6541
TAATAACTCA GTTCTTATTT GCACCTACTT CAGTGGACAC TGAATTTGGA AGGTGGAGGA 6601
TTTTGTTTTT TTCTTTTAAG ATCTGGGCAT CTTTTGAATC TACCCTTCAA GTATTAAGAG 6661
ACAGACTGTG AGCCTAGCAG GGCAGATCTT GTCCACCGTG TGTCTTCTTC TGCACGAGAC 6721
TTTGAGGCTG TCAGAGCGCT TTTTGCGTGG TTGCTCCCGC AAGTTTCCTT CTCTGGAGCT 6781
TCCCGCAGGT GGGCAGCTAG CTGCAGCGAC TACCGCATCA TCACAGCCTG TTGAACTCTT 6841
CTGAGCAAGA GAAGGGGAGG CGGGGTAAGG GAAGTAGGTG GAAGATTCAG CCAAGCTCAA 6901
GGATG
```

Fig. 8C

```
                     CA GGCCCCACAA AACCTAGATC TGCCCCAGTA TAACTAAATC 1501
TGGGACCATT TATTGAGCAA TTATTATGTG CCAAGTATTG CGCTGAGTGC TTCCAGAGCA 1561
TTATCTCCTT TAACCCCAGC ATAGTATGTC AGATGCTGTT TTACAGATGA GCCAACTGAG 1621
ACCAGAGATG CTCAGTCACT TGCCCAAGGT GACATGACTG ATATGGAATA GAGTCAAGAT 1681
TTTTTTTTTT TTTTTGACA CGGAGTCTCA CTCTGTCTCC CAGGCTGGAG TGCAGAGGCG 1741
CAATCTCAGC TCACTGCAAG CTCTGCCTCC CAGGTTCACG CATTCTCCTG CCTCAGCCTC 1801
CTGAGTAGCT GGGACTACAG GCACCCGCCA CCACACCTGG CTAATTTTTT GTATTTTTAG 1861
CAGAGACAGG GTTTCACCGT GTTAGCCAGG ATGGTCTCGA TCTCCTGACC TCGTGATCTG 1921
CCTGCCTCGG CCTCCCAAAG TGATGGAATT ACAGGTGTGA GCCACCGCGA CTGGCCAGAT 1981
TCAAGATTTG AACCCAGGTC CTCTTGGTCC CAGAGGCCCC TGTTTCTCAA CTCCCTAGCA 2041
TGCATACGCA CCTGTCCCTC TAGAGGTGCC TGCTTAAGTG TGCTCAGCAC ATGGAAGCAA 2101
GTTAGAAATG CTAGGTATAC CTGTAAAGAG GTGTGGGAGA TGGGGGGGAG GGAAGAGAGA 2161
AAGAGATGCT GGTGTCCTTC ATTCTCCAGT CCCTGATAGG TGCCTTTGAT CCCTTCTTGA 2221
CCAGTATAGC TGCATTCTTG GCTGGGCAT TCCAACTAGA ACTGCCAAAT TTAGCACATA 2281
AAAATAAGGA GGCCCAGTTA AATTTGAATT TCAGATAAAC AATGAATAAT TTGTTAGTAT 2341
AAATATGTCC CATGCAATAT CTTGTTGAAA TTAAAAAAAA AAAAAAAGT CTTCCTTCCA 2401
TCCCCACCCC TACCACTAGG CCTAAGGAAT AGGGTCAGGG GCTCCAAATA GAATGTGGTT 2461
GAGAAGTGGA ATTAAGCAGG CTAATAGAAG GCAAGGGGCA AAGAAGAAAC CTTGAATGCA 2521
TTGGGTGCTG GGTGCCTCCT TAAATAAGCA AGAAGGGTGC ATTTTGAAGA ATTGAGATAG 2581
AAGTCTTTTT GGGCTGGGTG CAGTTGCTCG TGGTTGTAAT TCCAGCACTT TGGGAGGCTG 2641
AGGCGGGAGG ATCACCTGAG CTTGGGAGTT CAAGACCAGC CTCACCAACG TGGAGAAACC 2701
CTGTCTTTAC TAAAAATACA AAAAATTCAG CTGGTCATGG TGGCACATGC CTGTAATCCC 2761
AGCTGCTCGG GAGGCTGAGG CAGGAGAATC ACTTGAACCA GGGAGGCAGA GGTTGTGGTG 2821
AGCAGAGATC GCGCCATTGC TCTCCAGCCT GGGCAACAAG AGCAAAGTT CGTTTAAAAA 2881
AAAAAAAAAG TCCTTTCGAT GTGACTGTCT CCTCCCAAAT TTGTAGACCC TCTTAAGATC 2941
ATGCTTTTCA GATACTTCAA AGATTCCAGA AGATATGCCC CGGGGGTCCT GGAAGCCACA 3001
AGGTAAACAC AACACATCCC CCTCCTTGAC TATCAATTTT ACTAGAGGAT GTGGTGGGAA 3061
AACCATTATT TGATATTAAA ACAATAGGCT TGGGATGGAG TAGGATGCAA GCTCCCCAGG 3121
AAGTTAGATA ACTGAGACTT AAAGGGTGTT AAGAGTGGCA GCCTAGGGAA ATTTATCCCG 3181
GACTCCGGGG GAGGGGGCAG AGTCACCAGC CTCTGCATTT AGGGATTCTC CGAGGAAAAG 3241
TGTGAGAACG GCTGCAGGCA ACCCAGGCGT CCCGGCGCTA GGAGGGACGA CCCAGGCCTG 3301
CGCGAAGAGA GGGAGAAAGT GAAGCTGGGA GTTGCCGACT CCCAGACTTC GTTGGAATGC 3361
AGTTGGAGGG GGCGAGCTGG GAGCGCGCTT GCTCCCAATC ACCGGAGAAG GAGGAGGTGG 3421
AGGAGGAGGG CTGCTTGAGG AAGTATAAGA ATGAAGTTGT GAAGCTGAGA TTCCCCTCCA 3481
TTGGGACCGG AGAAACCAGG GGAGCCCCCC GGGCAGCCGC GCGCCCCTTC CCACGGGGCC 3541
CTTTACTGCG CCGCGCGCCC GGCCCCACC CCTCGCAGCA CCCCGCGCCC CGCGCCCTCC 3601
CAGCCGGGTC CAGCCGGAGC CATGG
```

Fig. 9